

Dynamic qualitative filtering for linked (geo)visualisations: a prototype for exploring the collective memory of Instagram posts

Jones C.E., Guido D. and Severo. M².

¹University of Luxembourg | email: Catherine.jones@uni.lu or Daniele.guido@uni.lu

²Université Paris Ouest Nanterre La Défense

Summary

This paper describes a prototype geo extension to an open source graph-based tool designed for analysing co-occurrence in large multimedia collections of historical resources. Employing an interdisciplinary approach, combining tools and practices from humanities and geography, we explore how it can be extended to facilitate both social and spatial enquiry. For the prototype development we use Instagram resources that reference the Via Francigena – a significant cultural route connecting Canterbury, UK and Rome, Italy. The resulting tool facilitates dynamic qualitative filtering and multilevel views to explore the social-spatial collective memory of this cultural route.

KEYWORDS: Qualitative GIS, Instagram, collective memory, social-spatial visualisation

1. Introduction

Traditionally geo-visualisation environments are complex, visually cluttered and in the hands of the experts – they are not technology environments designed for use in the humanities. Furthermore, they are, in the main, based on quantitative data. For humanities scholars, digital historical resources constitute many diverse multimedia data, including photographs and audiovisual material, but the field is still dominated by text. Since we are now surrounded by an abundance of historical digital resources in the digital age (Fickers 2012), user-contributed datasets can offer an interesting perspective since they create qualitative spatial knowledge (Elwood 2009; Elwood and Mitchell 2015). Historians and social scientists have access to both traditional archival source material and/or volunteered social media data, but they require technology and interfaces that more appropriately meet their needs (Druker, 2013). Interfaces that encourage interpretation and reflection through close and distant reading and combine both social and spatial perspectives are now possible since the emergence of graph-based databases, web-based mapping solutions, APIs and the like. Furthermore, since Instagram represents mediated user-generated memories in the form of qualitative data (images, titles and tags), it is worthwhile using such interfaces to explore how these fragmented digital traces can shape a dataset for collective memory making.

2. Case Study – Via Francigena and collective memory

We describe a preliminary explorative study, employing an interdisciplinary approach and combining tools and practices from humanities and geography for social and spatial enquiry. The use case for the evaluation of the prototype was created based on the project “*Les espaces du patrimoine culturel numérique: topologies et topographies des itinéraires culturels*” (www.itinerairesculturels.fr) by Dr Marta Savero. The broader project explores new visualisation for socio-cultural perceptions and collective memory of cultural routes. Instagram posts (almost 9,000) referencing the Via Francigena – a significant cultural route connecting Canterbury, UK and Rome, Italy – were used as our dataset. The resulting prototype described in this paper facilitated dynamic qualitative filtering on multiple

linked views to explore social-spatial collective memory.

1.1. Feature requirements

We started by defining the requirements for the adaptation of the tool; this was achieved with a series of experiments conducted with CartoDB. The results of the experiments noted that the following features would enable an in-depth investigation of social-spatial collective memory:

- Ability to refine themes dynamically based on map windows – i.e. the map window is used as a search interface to define the area of interest and resulting visualisation
 - Applicable to spatial extent – if the user pans the density maps and co-occurrence maps are recalculated
 - Applicable to spatial scale – scale-dependent visualisations: if the user changes the zoom level the density maps and co-occurrence maps are recalculated
- Ability to display concentrations of tag use as density heat maps change
- Option of dynamically updating maps based on selected filter tags and displaying the related tags
- Possibility of dynamically linking the results to the resource view

1.2. Tool overview

The starting point of the research was to sort and organise the Instagram collection. To do this we integrated the data into the open source histogram tool (www.histogram.cvce.eu) which we adapted to include a geo-framework for powering interactive maps and qualitative filtering. histogram is an open source tool that was originally developed to support graph-based data exploration. It was initially designed to enable the investigation of co-occurrence networks in large diverse multimedia collections of historical resources such as treaties, meeting minutes, letters, photographs, etc. (Düring et al, 2015). It works with the textual components of each of the historical resources, for example the titles and captions (short descriptions of each resource) and carries out resource enrichment using Named Entity Recognition and Disambiguation (NERD) by calling external services to identify people, places and time periods that are mentioned in the resource. The co-occurrence network is then created, building up the connections between the different entities mentioned in the textual attributes of the resource. For our case study we are interested in the development of a prototype to explore geographical patterns of co-occurrence networks so that we can explore how these social media resources can inform our understanding of collective memory. Each Instagram post was therefore organised as a resource comprising of the structural attributes described in table 1.

Table 1 Structure of a resource ingested into (geo)histograph

Attribute	Element of the Instagram post	Description
Person	User ID	denoted with @ or persons mentioned in hashtags such as saints.
Place/ Location	Location picture was taken Places mentioned in resource context	Geotagged at time of posting Places referred to in the title or the hashtags
Theme	Hashtag	All text denoted with a # can be people, places or organisations
Object Title	Picture posted to instagram Caption included with post	Free text field where users can add a short title to their picture

1.3. Tool architecture

Figure 1 illustrates a simplified architecture of the histogram with its geo component. At the heart of the tool is a Neo4J graph database which stores the Instagram resources together with their attributes (table 1). A series of text analyses are undertaken, identifying locations, people and themes mentioned in the tags. A discovery script is used to analyse each text according to its language; in our resource dataset Italian, English and French are used. The Text Razor API returns a set of candidate entries for people, location and theme. For those entities encoded as locations the script then calls the API of Geonames and Google Geocoding to obtain the latitude and longitude of the location. If the two geo APIs are in consensus then the candidate entity is accepted and assigned. In a future version of the tool we expect to streamline the process by using DBpedia Spotlight to look up the coordinates of the locations returned by TextRazor and only calling the geo APIs for any locations that are missing. Once the discovery script has completed its processing and the resources have been enriched, the co-occurrence script sets out to generate all the links. On the server side, it builds the node-link relationship diagram, using the Jacard Similarity measure. The result is a social graph of co-occurrence for people, themes and locations in our resources.

(geo)histograph then uses its internal histogram API service to serve all the data as JSON files to the client side and in doing so builds the resource gallery, the graph interface and the prototype mapping interface. For the mapping interface the service translates the JSON into a geojson of geometry type point. Mapbox is then used to dynamically create the density maps and serve the vector tiles for the base map. Leaflet JS is then layered on top to create a simple map interface. A timeline component is created using D3.

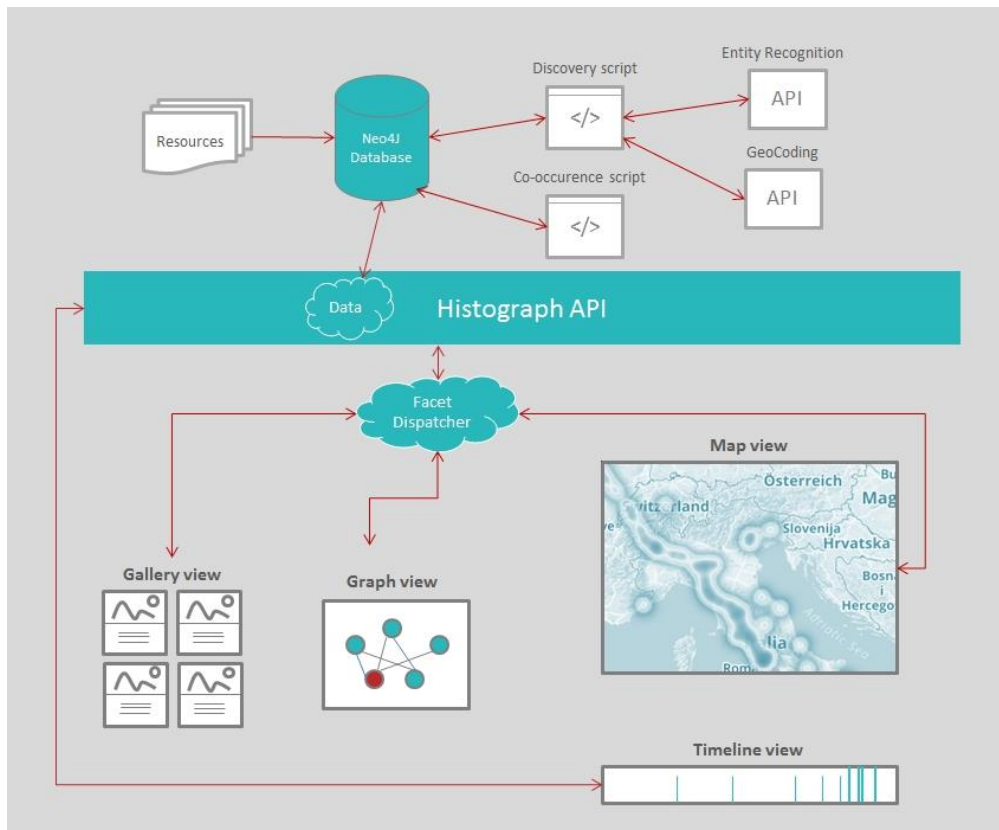


Figure 1 Simplified outline of (geo)histograph architecture

The histogram API service also includes a filtering function which is embedded into the mapping

interface. This facilitates dynamic co-occurrence text based filtering according to (1) if the map zoom or the spatial extent of the map window changes and/or (2) if the user selects one or more of the filter facets associated with a theme. If no facets are selected then the co-occurrence graph that was created on the server is used to generate the node-link diagram and the density map. If the filters are activated then the histogram API service calculates the co-occurrences on the fly.

3. Early results

The prototype has the ability to display the geocoded Instagram resources and filter the dataset based on entity type. The result is a tool composed of dynamically linked views which encourage seamless social-spatial investigation through the exploration of qualitative facets. Data are filtered and the resulting density maps are displayed according to the following qualitative dimensions: (1) the list of themes; (2) the resource type (e.g. Instagram or Twitter); (3) places where resources were originally geotagged; (4) places identified from the entity recognition of the hashtags (themes) and (5) people. It is also possible to seamlessly filter the data and recalculate the co-occurrence graphs based on the geographical extent of the map window and its scale. For each filter that is activated a unique URL is created and it is possible to activate as many filters as required.

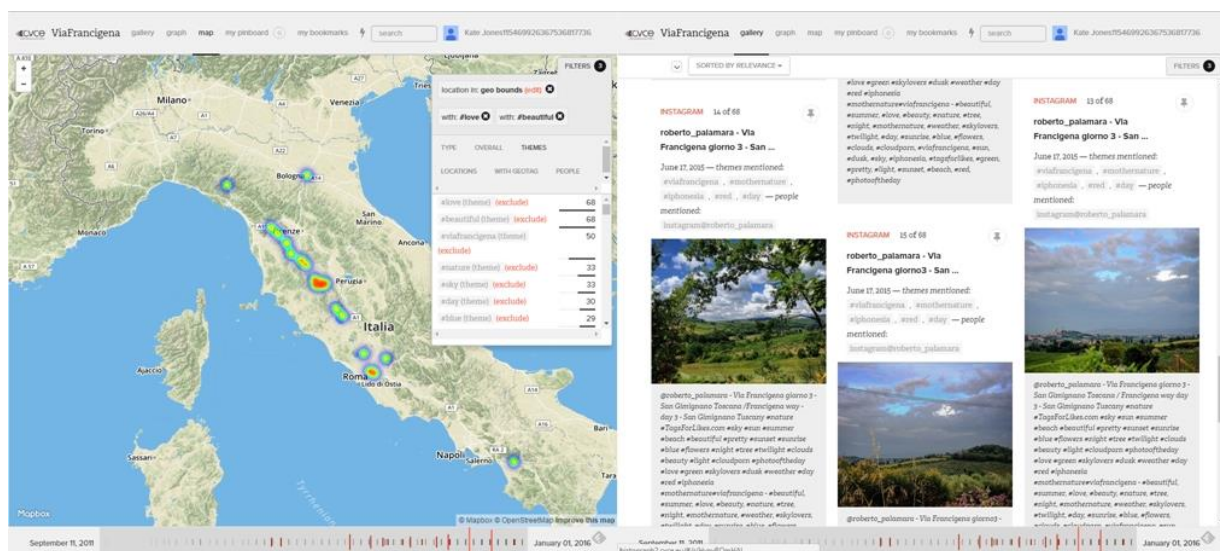


Figure 2 Simplified outline of (geo)histograph architecture

Figure 2 (left) highlights an example. The map shows Italy, and the data are filtered according to the themes of love and beauty. We see hotspots along the Via Francigena between Rome and Florence and that co-occurring themes are related to sky, nature, blue, sun and sunset. This type of dynamic filtering powered by co-occurrence graphs offers more flexibility to a geohumanities researcher than traditional spatial databases. By displaying the co-occurrences within the interface we can gather more insights about our historical resources. The interface uses a hot-cold colour scheme to imply higher concentrations of resources in red and lower concentrations in blue, although we expect to improve the cartography in future versions of the tool.

We can surmise that for the selected map area, the users of Instagram associate love and beauty with natural phenomena. A close reading of the resource gallery (figure 2: right) reinforces this perspective. The tools aid both close and far reading of the resources through dynamic visualisations that allow exploration of various social-spatial scales and visualisations.

4. Conclusion

The linked visualisations with interactive qualitative filtering provided through the histogram API offer a social-spatial scaled approach to close and distant readings of Instagram resources. Consequently we are able to conduct dynamic exploration through qualitative filtering to produce different visualisations, helping us to unravel a virtual collective memory.

5. Acknowledgements

This geo-component of this tool builds on the histogram tool, developed by the former CVCE Digital Humanities Lab (now University of Luxembourg) by Lars Wieneke, Daniele Guido, and Marten Düring. HistoGraph builds on a demonstrator application developed for the EC funded collaborative project CUBRIK (Grant agreement number 287704, 2011-14).

6. Biography

Catherine (Kate) is interested in the development of useful and usable tools for social-spatial and historical data analysis to enable meaningful narrative making. She has unique interdisciplinary perspective combining experience with GIS and Digital Humanities to explore spatial history. She joined the University of Luxembourg in July 2016 following a few years as the Digital Humanities Lab coordinator at the [CVCE](#) also in Luxembourg working on projects associated with ePublications, digital editions and user engagement. She is a co-investigator for the [H2020 CrossCult](#) Project – working on a pilot 4 to develop a geo-located game for reflective history. She studied for an MSc in Geographical Information Systems at University College London in 2002 and went on to complete a Knowledge Transfer Partnership, PhD and Post Doc at the same university, with a focus on interdisciplinary research for social and spatial data analysis now and in the past. This led her to the University of Portsmouth where as a Senior Lecturer in Human Geography where she was project director for the successful www.bombsight.org project.

Daniele is a designer and full-stack developer specialising in data visualisation and digital methods with a hybrid background in Fine Art (Fine Art Brera Academy, Milan, 2006) and Design (MSc in Communication Design, Politecnico di Milano, 2009). He spent several years at the [medialab Sciences Po](#) in Paris, where he helped to develop experimental web applications and tools applied to text analysis, data visualisation, collaborative writing and enhanced scholarly publications for social sciences, among them the front-end interface of the EU-funded project AIME (the AIME platform modesofexistence.org), also in collaboration with the designers of the [Density Design](#) Lab team, Politecnico di Milano. He joined the [CVCE Digital Humanities Lab team](#) in March 2015, where he was responsible for the design and development of [histograph](#), an open source platform for the graph-based exploration and crowd-based indexation of multimedia collections. As a member of the DH Lab he experimented with new and emerging technologies for graph databases, real-time interactions in web applications, named entity extraction and disambiguation tools, network visualisation, geographic visualisation and text analysis. As part of the DEIS team following the CVCE's integration into the University of Luxembourg he designed paper and mobile prototypes for a mobile game in the built environment in the context of [CrossCult](#). He is now in responsible for the design and the development of the web platform for enhanced publications in connection with the [RESuME](#) project.

References

Düring, Marten, Lars Wieneke, and Vincenzo Croce. 2015. Interactive Networks for Digital Cultural Heritage Collections - Scoping the Future of HistoGraph. In *Engineering the Web in the Big Data Era*, edited by Philipp Cimiano, Flavius Frasinicar, Geert-Jan Houben, and Daniel Schwabe, 613–16. Lecture Notes in Computer Science 9114. Springer International Publishing.

- Elwood, Sarah. 2009. 'Geographic Information Science: New Geovisualization Technologies — Emerging Questions and Linkages with GIScience Research'. *Progress in Human Geography* 33 (2): 256–63. doi:10.1177/0309132508094076.
- Elwood, Sarah, and Katharyne Mitchell. 2015. 'Technology, Memory, and Collective Knowing'. *Cultural Geographies* 22 (1): 147–54. doi:10.1177/1474474014556062.
- Fickers, Andreas. 2012. Towards A New Digital Historicism? Doing History In The Age Of Abundance. *VIEW Journal of European Television History and Culture*. [online] <http://orbilu.uni.lu/handle/10993/7615>.
- Goodchild, Michael F. 2007. 'Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0'. *International Journal of Spatial Data Infrastructures Research*, 24–32. 'Goodchild_VGI2007.pdf'. 2016. Accessed September 19. http://ncgia.ucsb.edu/projects/vgi/docs/position/Goodchild_VGI2007.pdf.